Biocybernetics and Biomedical Engineering
2010, Volume 30, Number 4, pp. 29–47

Analysis of Extracted Cardiotocographic Signal Features to Improve Automated Prediction of Fetal Outcome

MICHAŁ JEŻEWSKI[1], ROBERT CZABAŃSKI[1], JANUSZ WRÓBEL[2,*], KRZYSZTOF HOROBA[2]

[1] *Division of Biomedical Electronics, Institute of Electronics, Silesian University of Technology, Gliwice, Poland*
[2] *Department of Biomedical Signal Processing, Institute of Medical Technology and Equipment, Zabrze, Poland*

Cardiotocographic monitoring based on automated analysis of the fetal heart rate (FHR) signal is widely used for fetal assessment. However, the conclusion generation system is still needed to improve the abnormal fetal outcome prediction. Classification of the signals according to the predicted fetal outcome by means of neural networks is presented in this paper. Multi-layer perceptron neural networks were learned through seventeen time-domain signal features extracted during computerized analysis of 749 traces from 103 patients. The analysis included estimation of the FHR baseline, detection of acceleration and deceleration patterns as well as measurement of the instantaneous FHR variability. All the traces were retrospectively verified by the real fetal outcome defined by newborn delivery data. Influence of numerical and categorical representation of the input signal features, different data sets during learning, and gestational age as additional information, were investigated. We achieved the best sensitivity and specificity for the neural networks fed with numerical input variables together with additional information on the gestational age in the categorical form.

K e y w o r d s: cardiotocography, fetal heart rate monitoring, fetal outcome prediction, pattern classification, signal analysis

1. Introduction

Cardiotocographic (CTG) monitoring is a routine procedure for assessment of the fetal state during pregnancy and labour. It relies upon non-invasive recording of Fetal Heart Rate (FHR), maternal Uterine Contractions (UC) and fetal movement activity.

* Correspondence to: Janusz Wróbel, Department of Biomedical Signal Processing, Institute of Medical Technology and Equipment, ul. Roosevelta 118, 41-800 Zabrze, Poland, e-mail: januszw@itam.zabrze.pl
Received 11 February 2010; accepted 09 July 2010

Early detection of the fetus at risk helps to avoid dangerous situations which could be more difficult or even impossible to manage in the newborn. The CTG signals undergo analysis aimed at extraction and quantitative description of the features essential for classification of the traces as corresponding to normal or abnormal fetal state. When classifying the FHR signals, the baseline, variability and the presence of acceleration/deceleration patterns have to be assessed. The resting level of the fetal heart rate (the basal FHR) between 120 and 160 bpm is the fundamental pattern. Accelerations of the FHR as temporary increases of fetal heart rate in response to fetal movements are a sign of fetal central nervous system alertness and fetal well-being. The temporary decreases of the FHR called decelerations usually reflect such risky events as compression of the umbilical cord.

At present, quantitative analysis of CTG records is performed with a help of computerized fetal monitoring system. Its tasks are: analysis of data incoming from bedside monitors, dynamic presentation of signals along with analysis results, as well as storing and printing the data [1]. The system ensures easy and fast access to archived records and convenient following up their longitudinal changes. Automated analysis of the cardiotocographic signals is able to extract all the features that are hidden for visual evaluation done by clinicians. It is very important especially for the determination of the beat-to-beat FHR variability, which is crucial for the fetal state assessment. Additionally, stable computer algorithms and threshold values significantly increase repeatability and objectivity of signals analysis. However, fetal assessment is still done by clinicians who finally classify the trace features as relating to normal or abnormal fetal outcome. Such prediction of the fetal outcome during pregnancy is possible, because in perinatology it is assumed that the fetal state can not change rapidly. In other words, a newborn whose state just after delivery has been evaluated as normal had to have developed properly, excluding of course acute complications during labour. It was found that visual interpretation of CTG traces is characterized by low inter- and intraobserver agreement, which may lead to erroneous diagnosis. The benefit of fetal monitoring is that the reassuring CTG features are usually confirmed by normal fetal outcome. While the abnormal signal patterns can relate both to abnormal and normal real fetal state, and false assessment of the abnormal fetal state very often causes unnecessary operative interventions.

Since the cardiotocography is a primary method for fetal state assessment, looking for automated methods for efficient conclusion generation is extremely needed. Possibility of handling of complex data sets, capability of learning and generalization, and distributed pattern recognition process, make a use of neural networks particularly attractive for medical application. In obstetrics it concerns mainly fetal outcome assessment [2], prediction of preterm birth [3], low birth weight [4] or even newborn gender [5]. In the learning process of the neural networks, aimed at cardiotocographic traces classification, a clinical experts'

knowledge is applied, which is based on evaluation of selected parameters from the newborn description [6]. The input set for automated classifier was usually formed by time and/or frequency domain parameters from computerized analysis of the FHR signal [7–9]. Sometimes these parameters were converted into new artificial features by means of grammatical evolution [10]. The additional features were based on discrete wavelet transform [11] or approximate entropy [12]. Different combinations were used as input data and usually the initial number of inputs was reduced to ensure better classification performance. The raw FHR and UC values averaged over two-minute intervals were used in [13], where the neural network was trained to classify CTG into three categories: physiological, suspicious and pathological, the neural network reached the same pathology prediction as the expert – in 69.7% of cases. More often, the raw signals were applied when the neural network task was to extract and classify clinically important features: the FHR baseline, acceleration and deceleration patterns [14–16]. Both raw signals and seventeen automatically extracted features were the inputs of the neural networks trained to interpret the CTG records using clinical experts and the non-stress test rules [17]. Generally, the Multi-Layer Perceptron (MLP) was applied more often than other classifiers tested: classical statistical methods (linear, quadratic and logistic discriminant analysis) [12], Neuro-Fuzzy Inference System [18, 19], Artificial Neural Network Based on Logical Interpretation of fuzzy if-then Rules [20] or self-organizing (Kohonen) maps [21]. Retrospective verification of CTG records classification by means of fetal outcome was applied only in [22].

Concluding, usually the automated approach provided better results than the human experts. However, particular emphasize was placed on the automated classification method of the CTG traces itself, whereas the different aspects of input data applied during the learning process should have been considered as well. First of all, the fetal trace classification is based on automatically extracted signal features. It means that the quality of the algorithms applied in quantitative analysis of monitoring records is extremely crucial for the prediction of the fetal outcome as being normal or abnormal. Thus, in this work we focused on the CTG signal features extraction and testing of different approaches to input data preprocessing. We investigated how two possible representations of the input variables – numerical versus categorical, the different structures of the learning data sets as well as additional information on a fetal gestational age (at which the fetal signals were registered) affect the quality of the CTG signals classification. In the proposed work, the learning process was accomplished with the true fetal outcome evaluated on the newborn data by the experts just after delivery. Several experiments were performed using multi-layer perceptron neural networks as the most representative since they have been most often used. They were aimed to show the directions in which the performance measures – the prognostic indices – can be expected to change. Therefore, the obtained results do not represent their maximum values.

2. Material and Methods

2.1. Data Collection

The data were obtained from the archive of computerized fetal surveillance system MONAKO [1]. They consisted of a set of parameters of quantitative description of the CTG signals in time domain together with the associated medical history referring to the patients and their newborns. The CTG signals were recorded using HP series 50 monitors. The FHR was acquired via a pulsed Doppler ultrasound transducer placed on the maternal abdomen. Uterine contractions were recorded via a strain gauge transducer. The monitor provided every 250 ms the consecutive digital measurements of both signals, the resolution was 0.25 bpm for the FHR and 0.5 of relative unit for the uterine activity. We rejected records from patients with incomplete delivery and newborn data forms, as well as like in [23], those with fetal malformations diagnosed before or after delivery, multiple pregnancy and acute complications during labour (like difficult fetal extractions or anaesthesia complications). Finally, we obtained 749 records from 103 patients, where 210 (28%) records related to abnormal fetal outcome. The number of traces recorded from particular patient varied from one to ten, and they were acquired between 28th and 42nd week of gestation.

2.2. Features Extraction

Fetal heart activity is described by changes of the cardiac intervals T_i determined between two consecutive heart beats or using the instantaneous fetal heart rate values FHR_i, which is an extrapolation of the interval T_i into one-minute period:

$$FHR_i = \frac{60000}{T_i\,[miliseconds]} \quad [beats\ per\ minute] \tag{1}$$

Both the representations are used interchangeably during computerized analysis. Significant FHR variability in time (Fig. 1) is caused by complex heartbeat regulation system. There are a number of different variability patterns which can be grouped into:

– changes of the basal fetal heart rate called baseline that comprise very slow and usually long-lasting decrease or increase of the heart rate, which when exceeding the established thresholds, are defined as bradycardia and tachycardia respectively;

– changes of the fetal heart rate in certain direction, e.g. transitory increases above the baseline defined as accelerations of the FHR, as well as transient slowing of the FHR in relation to the baseline called decelerations;

– short-lasting changes of the FHR also called instantaneous variability. There are two types of these variability: short-term variability with changes of consecutive T_i intervals duration (called beat-to-beat variability), and long-term variability with periodical changes of beat-to-beat variability concerning both direction and magnitude (called oscillations of FHR).

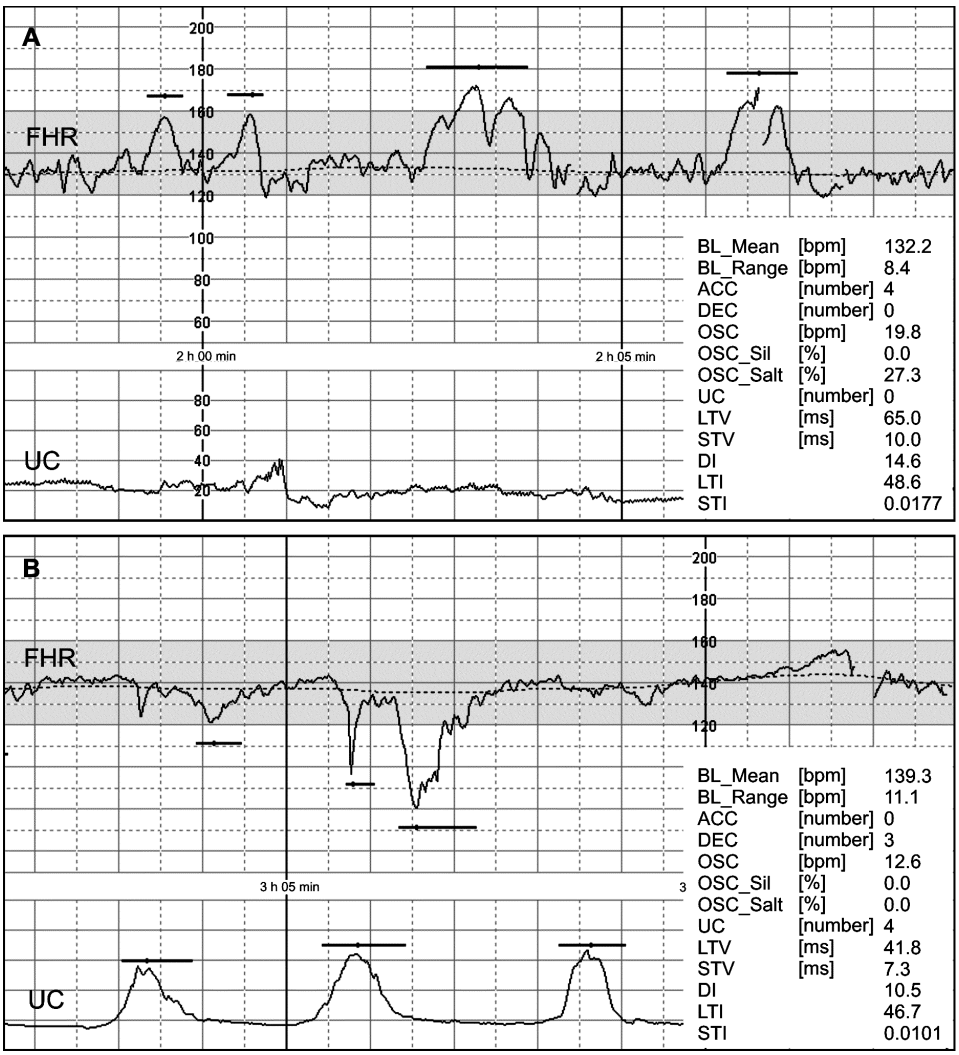

Fig. 1. Cardiotocographic signals. Two screens with cardiotocographic signals: A – accelerative and B – decelerative. The horizontal bars above the waveforms identify the recognized trace patterns. Additional window presents thirteen signal parameters: ACC, DEC and UC state the number of patterns recognized in segment, the OSC_Sil, OSC_Salt – percentage of duration of a given oscillation episode in the segment duration and the rest ones represent mean values of the given feature in the segment (see Table 1)

2.2.1. Preprocessing

In computerized signal analysis the instantaneous FHR variability is described by parameters which require the signal to be in a form of time event series – consecutive T_i intervals, and/or as evenly sampled signal. The bedside monitors provide the FHR measurements every 250 ms, which prevents from loss of short cardiac intervals.

However, long intervals can be represented by up to five values. Therefore, the first step of the analysis is extraction of the series of the consecutive cardiac intervals. Simple solution of this problem is to replace samples of the same value with one valid to mark others as duplicated ones [24]. To improve this approach, the authors developed a novel algorithm for identification and removing the duplicated samples [25]. It relies on a formula defining the range of the number of real intervals x represented by a series of n samples of the same value T_i and occurring every $t = 250$ ms period:

$$\frac{(n-1)\cdot t + T_i - T_{i+1}}{T_i} < x < \frac{(n+1)\cdot t + T_i - T_{i+1}}{T_i} \tag{2}$$

If (2) have only one solution (as an integer), it is the exact number of intervals. In case of two solutions obtained, the probability criteria are used to choose the one, with possible error of ± 1 interval for each particular series of duplicated samples. Because such doubtful cases are very rare accuracy of the extraction of the event series is very high.

2.2.2. Artefacts Removal

In commonly used criteria [26] to control the instantaneous changes of T_i, too wide range of acceptability is applied, mainly for correct recognition of the slopes of acceleration and deceleration pattern. The criteria are efficient enough for signal displaying or printing, but in many cases the values assumed as correct distort the analysis of beat-to-beat variability. Therefore, we proposed more precise criteria for T_i interval validation [25]. In the first step the signal measurements are accepted only when they fulfil the condition:

$$T_{i-1} - 0.10\cdot\Delta_{i-1} < T_i < T_{i-1} + 0.15\cdot\Delta_{i-1} \tag{3}$$

where:

$$\Delta_{i-1} = \begin{cases} T_{i-1} - 300ms & for \quad T_{i-1} \geq 320ms \\ 20ms & for \quad T_{i-1} < 320ms \end{cases} \tag{4}$$

The interval T_i is accepted if it belongs to the group of three consecutive intervals fulfilling (4). To remain the correct intervals within the slope of acceleration/deceleration patterns [27], the validation is carried out bidirectionally. A given interval is considered as incorrect only if it does not meet the criteria in both directions. The final verification of T_i intervals preliminary classified as incorrect is based on the analysis of direction of beat-to-beat changes. A given T_i interval is marked as incorrect if the derivative sign reverses for this interval

and a product of differences between a given interval and the neighbouring ones exceeds the value of 35.

2.2.3. Baseline

The baseline term has been introduced to define the changes in time of the so-called basal level of the fetal heart rate. According to the FIGO guidelines [28]: "the baseline is the mean level of the FHR when this is stable, accelerations and decelerations being absent". The FHR baseline and the acceleration/deceleration patterns are recognized as the most clinically important features of the FHR interpreted both in classical visual and in computer-aided analysis. The FHR baseline estimation is based on nonlinear filtering, where the filter settings are being adaptively changed using the amplitude of the input signal and some statistical parameters of the record. Before filtering the averaged FHR values are calculated over a moving window of 2.5 s width using only the correct interval values T_i. The filtering process is driven by prominent rate value F obtained from the frequency distribution of the fetal heart rate values in the analyzed record [29, 30]. When the input signal significantly differs from the actual prominent rate, the filtering stops and the constant value is kept on the filter output.

An iterative filtering of the FHR signal ensures a good FHR baseline fitting, both during small FHR changes and accelerations and decelerations episodes. The number of iterations applied to a given FHR segment increases as the deviation of this segment from the baseline increases. After each filtering the output signal is modified to restore the primary samples in these parts where the differences between their values and the baseline (the output signal after a given iteration step) are less then the established thresholds. With each consecutive iteration the thresholds are gradually decreased. Depending on the number of iterations the filter cut-off frequency varies from 0.126/min for one iteration to 0.053/min for five iterations. During the first 45 minutes of recording, the baseline is estimated every one minute for a whole signal and the previous baseline is replaced. As the time goes on, the baseline is always estimated over the last 45 minutes. To avoid discontinuity, the initial fragment (15 minutes) of the final baseline is obtained by weighted averaging of the previously and currently determined values. Two parameters directly describing the FHR baseline have been included in the input data set for neural network: the mean baseline value (*BL-Mean*) and the fluctuation range of baseline values calculated as a difference between maximum and minimum values (*BL-Range*).

2.2.4. Acceleration and Deceleration Patterns

The fetal heart rate patterns classified as acceleration and deceleration (A/D) episodes represent transient deviation of the FHR signal around the baseline with established range of amplitude and duration. According to the FIGO definition [28]

the acceleration is recognized if the increase in the FHR above the baseline is of 15 bpm or more and lasts 15 s or more. Deceleration is a transient episode of slowing of the heart rate below the baseline level of more than 15 bpm and lasting minimum of 10 s. The first stage in the recognition procedure is a preliminary detection of the FHR deviations regardless of the signal loss influence, using the signal completed with interpolated samples. In the next stages, the preliminary detected and classified deviations undergo verification regarding the influence of signal loss. Some interpolated samples are rejected within A/D episodes, that leads to shortening, splitting into separate episodes or complete rejection of the pattern. After the correction, the signal segments are classified as the acceleration and deceleration patterns only if they still fulfil the A/D definition thresholds and the percentage of valid averaged values left within them is below 30% and 50%, respectively. All the finally accepted accelerations and decelerations are analyzed to calculate their detailed parameters: duration, amplitude and area. Numbers of the A/D patterns detected per hour have been included into the neural network inputs set as *ACC* and *DEC* variables.

2.2.5. Instantaneous Variability

Analyzing the instantaneous variability of the fetal heart rate, the changes of duration of consecutive cardiac intervals T_i are defined as the short-term variability or alternatively the beat-to-beat variability. Periodical changes of the short-term variability leading to the fetal heart rate fluctuation in relation to their mean value are defined as the long-term variability. The FHR variability is quantitatively described by mathematical formulas based on various combinations of mean value, standard deviation or interquartile range [31]. Table 1 lists parameters used as the input data for the neural networks. All the variability indices are calculated for separated one-minute signal windows with samples averaged over 2.5 s. Thus, a given index, depending on the signal loss, can be calculated from up to $N = 24$ values. Exceptions are the indices marked with BB, which are calculated using the signal in a form of M events T_i within one-minute fragment. With exception of silent and saltatory oscillations, the final value of a given index relates to the whole signal record and it is calculated as a mean of all one-minute values.

2.2.6. Uterine Contraction Activity and Fetal Movements

The signal of uterine contraction (UC) activity from the fetal monitor is in a range of 0 to 100 relative units (the value of 10 units corresponds approximately to the strain of 100 grams). Contraction is represented on the UC curve as an increase above the so-called uterine basal tone – some basal strain exerted by the uterine muscle on the strain-gauge transducer when contractions do not occur. The algorithm for automated detection of contraction patterns implemented by authors [32] is based on the analysis

Table 1. Definitions of selected indices describing the fetal heart rate variability

Parameter	Description
STV STV-BB	Dawes short-term variability index: $STV = \frac{1}{K}\sum_{i=1}^{K-1}\lvert T_{i+1} - T_i\rvert \quad [ms]$
LTV	Dawes long-term variability index: $LTV = T_{max} - T_{min} \quad [ms]$, where: $T_{max} = \max\{T_1, \ldots T_K\}$, $T_{min} = \min\{T_1, \ldots T_K\}$
DI DI-BB	Yeh short-term variability index: $DI = \sqrt{\frac{1}{K-2}\sum_{i=1}^{K-1}\left(d_i - \bar{d}\right)^2} \quad [a.u.]$, where: $d_i = \frac{T_i - T_{i+1}}{T_i + T_{i+1}}$, $\bar{d} = \frac{1}{K-1}\sum_{i=1}^{K-1}$
LTI STI STI-BB	De Haan long-term variability index: $LTI = IQR\,(r_i)$ [*a.u.*] De Haan short-term variability index: $STI = IQR\,(\varphi_i)$ [*a.u.*] where: *IQR* – interquartile range, $r_i = \sqrt{T_{i-1}^2 + T_i^2}\,; i = \langle 1 \ldots K \rangle$, $\varphi_i = arctg\left(\frac{T_i}{T_{i-1}}\right); i = \langle 1 \ldots K \rangle$
OSC	Oscillation amplitude: $OSC = FHR_{max} - FHR_{min} \quad [bpm]$ where: $FHR_{max} = \max\{FHR_1, \ldots FHR_K\}$, $FHR_{min} = \min\{FHR_1, \ldots FHR_K\}$
OSC-Sil	Percentage of silent oscillation ($OSC \leq 5$ bpm) in a whole trace: $OSC - Sil = \frac{1}{L} \cdot \sum_{i=1}^{L} A_i \cdot 100 \quad [\%], \quad A_i = \begin{cases} 1 & OSC_i \leq 5bpm \\ 0 & otherwise \end{cases}$ *L* – number of minutes in the trace
OSC-Salt	Percentage of saltatory oscillation ($OSC \geq 25$ bpm) in a whole trace: $OSC - Salt = \frac{1}{L} \cdot \sum_{i=1}^{L} A_i \cdot 100 \quad [\%], \quad A_i = \begin{cases} 1 & OSC_i \geq 25bpm \\ 0 & otherwise \end{cases}$

$K = N$ – for the signal in a form of 2.5 s samples.
$K = M$ – for the signal in a form of event series T_i (marked as *BB*).

of frequency distribution of the UC values in the moving window of four-minute width. In each window the modal value is selected as a consecutive basal tone sample. The contraction starts if the UC value exceeds the level of 10 units above the basal and remains above the detection threshold longer than 30 s with amplitude more than 20 units. Uterine contraction pattern is described by its beginning, duration, amplitude and area under the curve. The number of contractions (*UC*) given per hour detected in signal was used as an input parameter for the neural network. The number of fetal movements (per hour) perceived by mother during the entire monitoring session was added to the neural network inputs set as *MOV* variable.

2.3. Fetal Outcome

For the period of pregnancy, the fetal state assessment, which is done by clinician basing on evaluation of CTG trace features, can not be verified. There is no other non-invasive diagnostic method that would be able to evaluate the fetal state with higher accuracy and thus to play a role of a reference. Fortunately, in perinatology it is assumed that the fetal state can not change rapidly. Thus, the real fetal state assessed after delivery (fetal outcome) can be retrospectively related to the previous recordings of a given patient. When using neural networks for classification of CTG signals, the learning process is accomplished with the known results of fetal outcome evaluated by experts just after delivery with a help of three main attributes of the newborn. The Apgar score is a simple method for evaluation of newborn's physical condition just after the childbirth using five factors: appearance, pulse, grimace, activity and respiration. Each factor is scored on a scale of 0 to 2, with 2 being the best score. The resulting sum – the Apgar score-ranges from 0 to 10, and the value below 7 is regarded as abnormal. Percentile of birth weight is determined basing on neonatal birth weight in relation to its reference value derived form Polish national data charts within the range from 28th to 44th completed week of gestation. The reference percentiles are stratified by infant sex, and the gestational age is calculated by USG. Birth weight below the 10th reference percentile is regarded as abnormal. At the time of birth the umbilical cord blood sampling for gas values analysis (especially for pH measurement) is considered as very important for fetal oxygenation status. The sample of blood from a clamped segment of umbilical cord (usually artery) is used. The value of pH below 7.20 means an abnormal fetal state.

The neural network output (classification result) represents the predicted fetal outcome, but in practice, it means the fetal state at the time of CTG monitoring. In our application the developed neural network has two-state output representing normal or abnormal fetal outcome. Common approach in clinical practice is to assume the fetal outcome as abnormal, if at least one attribute is outside the physiological range. In our research material we noted that the neonatal birth weight was the most decisive attribute as it classified 131 fetal outcomes as abnormal, whereas Apgar score 45 and pH only 17. The rest 17 fetal outcomes were classified as abnormal due to two or all three newborn attributes being outside their physiological ranges. The classifier output can be defined alternatively – as OR function of outputs of three neural networks predicting the outcome attributes separately. We compared these two approaches of neural network output realizations [33], and the obtained results showed lower efficiency of the classifier based on OR function of separate outputs.

2.4. Neural Networks Modelling

The set of 17 parameters of quantitative description of the CTG signals was chosen as input variables for the neural networks. The input data were normalized according to

their minimum and maximum values. Fifteen parameters describe the FHR features in time domain: the baseline (two indices), the number of recognized acceleration and deceleration patterns (two), the short-term variability (six), as well as the long-term variability parameters (five indices). Additionally, the number of identified uterine

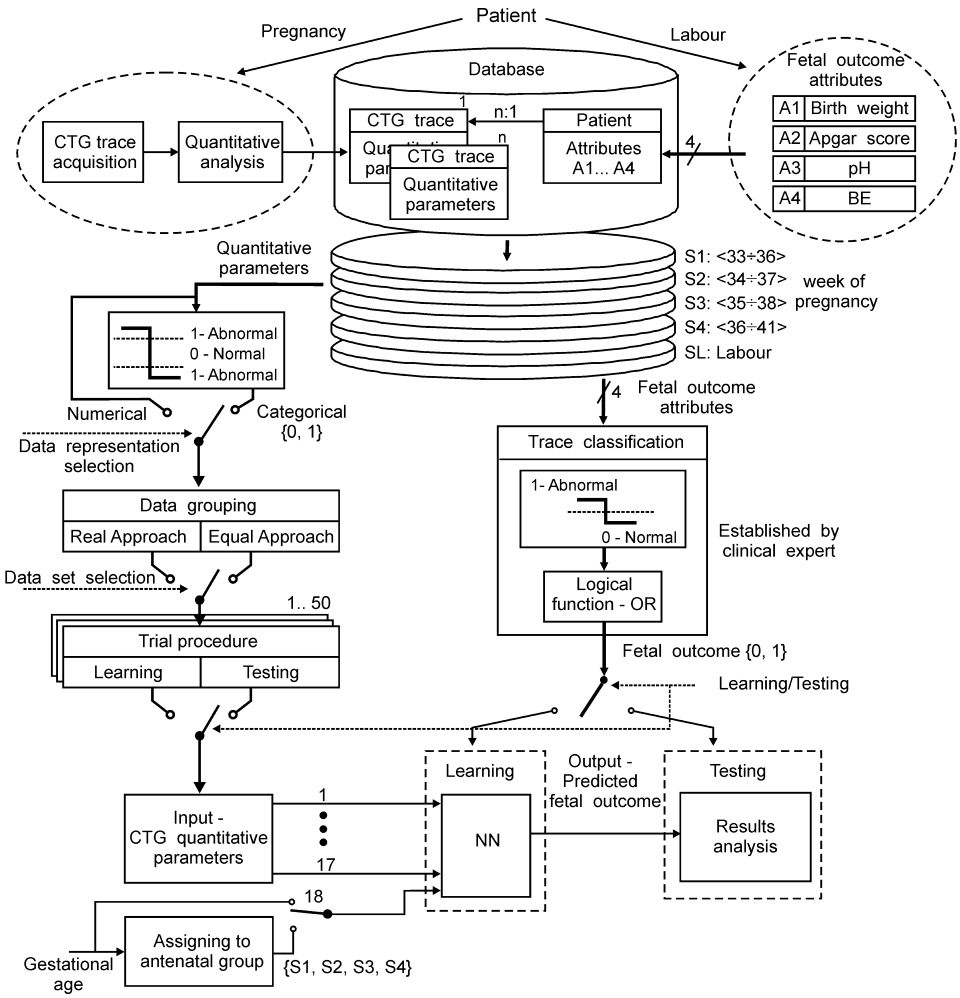

Fig. 2. General scheme of the experiments performed with application of the neural networks for the CTG trace classification. Seventeen quantitative parameters represent each signal record taken from the database. They are fed to neural network inputs in numerical and categorical form. Gestational age is used additionally either as the week number or the antenatal group number. During the learning stage the input data can be grouped in two ways: Real Approach – with the original sizes of classes maintained, and Equal Approach – with classes adjusted to the same size by removing some randomly selected records referring to the normal outcome. Fifty trials with randomly arranged contents of the data sets are applied. Classification efficiency is determined in relation to fetal outcome for a whole research material and separately in the particular antenatal groups

contractions and the number of fetal movements were involved. We used MLP neural networks with the sigmoid activation function. The number of neurons in a hidden layer were changed in two ranges: from 2 to 10 with one-neuron step and from 5 to 250 with five-neuron step. Additionally, different learning time in epochs of 200, 500, 1000, 5000, 10000 as well as the learning rate with values of 0.001, 0.01, 0.1, 0.15, 0.5 were applied. The momentum term was set to 0.3. The values of weights were initialized using a uniform random generator inside the interval [0.0, 1.0]. The output was a single neuron and the threshold between the two classes analyzed was automatically determined, while minimizing the classification error. The steepest descent gradient algorithm was used.

To avoid the situation when the network with a given structure provides very good results only by chance, we applied a set of trials with randomly arranged contents of data sets (Fig. 2). In every experiment, the cases were 50 times randomly assigned to three data sets: learning, validating and testing. As a result we obtained 50 neural networks with a given constant structure, but with different performance parameters resulted from the learning process. The normal and abnormal cases were partitioned in the learning, validating and testing data sets in proportion 50%, 25% and 25% respectively. The ratio of cases with normal fetal outcome to abnormal one in each set was constant in all trials. The exception was the Equal-Approach learning which will be described later. Results of the particular experiments are presented as mean values (with standard deviations) calculated for all trials.

Table 2. Summary statistics of the classification results obtained for three NNs designed with different data representations and learning data set organization

Data representation:	Categorical values	Numerical values	
Learning type:	Real-Approach	Real-Approach	Equal-Approach
Prognostic index (mean ± SD [%])			
SE	52.9 ± 12.4	64.7 ± 6.6	59.2 ± 7.1
SP	65.8 ± 11.6	66.6 ± 3.6	72.5 ± 3.2
PPV	37.9 ± 7.0	43.9 ± 5.3	39.0 ± 4.7
NPV	78.9 ± 4.0	82.3 ± 3.6	85.7 ± 2.5
CC	62.5 ± 7.5	66.0 ± 4.6	69.5 ± 4.5
OI	56.0 [a] ± 8.6	62.4 ± 6.7	61.6 ± 8.5

[a] – statistically significant difference on level $p < 0.001$.

2.5. Experiments

Pregnant woman can be monitored many times, especially in case of high-risk pregnancy, which causes that a given fetal outcome is related to several CTG traces. For the classification procedure it is possible to select one trace for each patient, for example the one registered as close as possible to the delivery. However, in [33] we stated that leaving the research material unchanged gives better results, and thus

this approach was used in the current study. The general scheme of the experiments carried out in this study is illustrated in Fig. 2.

In the database the fetal signal records are represented by 17 quantitative parameters which were fed to the neural network inputs in numerical and categorical form. Information on gestational age was used additionally either as the week number or the antenatal group number. During the learning stage the input data were grouped in two ways: Real Approach – where the original sizes of classes were maintained, and Equal Approach – with classes adjusted to the same size by removing some randomly selected records referring to the normal outcome. Fifty trials with randomly arranged contents of data sets were applied. Classification efficiency was determined in relation to the fetal outcome for a whole research material and separately for the particular antenatal groups. Eleven parameters of quantitative description of CTG signals could be converted from their original numerical values into categorical ones basing on the established ranges of physiology [34]. After the conversion the value of 0 means that the corresponding numerical value is within the normal range, whereas 1 means the abnormal value. The range is a function of gestational age at which the CTG signals were recorded. Descriptive statistics of the input parameters and its abnormal and normal values within the material collected are listed in Table 3. For both data set representations the neural networks were proposed.

Table 3. Descriptive statistics of input parameters ranked according to the importance index determined as the mean value of the numerical inputs positions taken in all fifty trials with a use of the MLP neural network

Input parameter		Abnormal range		Mean ± SD	Min ÷ Max	Mean value of ranks
		Number of traces	Percentage of traces [%]			
ACC	[number]	205	27.4	8.49 ± 5.77	0.0 ÷ 36.3	2.60
LTI	[–]	167	22.3	24.08 ± 7.79	5.8 ÷ 50.5	3.06
STI	[–]	138	18.4	1.14 ± 0.33	0.3 ÷ 2.6	4.64
STI-BB [a]	[–]	–	–	0.50 ± 0.11	0.2 ÷ 1.0	5.40
STV-BB [a]	[–]	–	–	2.86 ± 0.73	1.1 ÷ 6.7	5.74
DI-BB [a]	[–]	–	–	4.61 ± 1.19	1.6 ÷ 10.5	6.22
DI	[–]	188	25.1	9.34 ± 2.77	2.6 ÷ 19.9	8.76
UC	[number]	65	8.7	2.91 ± 4.97	0.0 ÷ 27.1	9.02
LTV	[ms]	117	15.6	42.45 ± 11.04	10.8 ÷ 75.2	9.48
BL-Range [a]	[bpm]	–	–	12.54 ± 2.61	0.0 ÷ 21.4	11.24
BL-Mean [a]	[bpm]	–	–	143.62 ± 8.94	115.4 ÷ 72.4	11.40
MOV	[number]	423	56.5	33.38 ± 47.54	0.0 ÷ 506.2	11.76
DEC	[number]	30	4.0	1.56 ± 2.87	0.0 ÷ 25.0	11.92
OSC-Sil	[%]	206	27.5	6.01 ± 8.57	0.0 ÷ 70.3	12.58
OSC [a]	[bpm]	–	–	14.58 ± 3.79	5.0 ÷ 26.0	12.60
STV	[ms]	137	18.3	6.03 ± 1.86	1.4 ÷ 13.4	13.24
OSC-Salt	[%]	15	2.0	11.44 ± 0.81	0.0 ÷ 75.0	13.34

[a] – reference chart not determined for this parameter.

In highly-developed countries the fetal outcome is normal in most of the cases. Our relatively high number of abnormal cases – 210 (28%) in relation to the normal ones – 539 (72%), is caused by the fact that the research material was obtained from clinical centre which represents in Poland the highest level of perinatal care. In such centres the number of abnormal cases can reach up to 40%. However, research material with still extreme difference in class sizes in the learning process may be considered as unfavourable. On the other hand, class sizes corresponding to the real distribution should give better performance of neural networks. Therefore, during the learning stage the input data were grouped in two ways (Fig. 2), and the results of classification were compared. The first grouping was called the Real-Approach, in which the original sizes of classes were maintained in the learning data set. Since we partitioned the normal and abnormal cases into the learning, validating and testing data sets with proportion 50%, 25% and 25% respectively, the learning set have included initially 370 cases referring to normal and 105 corresponding to abnormal fetal state. In the second data grouping called the Equal-Approach, both classes were equalized by removing some randomly selected records referring to the normal outcome. Finally, the reduced learning data set comprised the same number (105) of the normal and abnormal cases.

A large number of the input parameters with quantitative description of the CTG signals leads to question about their real influence on the classification quality. This was investigated by estimation of an importance index for each particular input variable within all 50 learning trials. The importance index was defined as the ratio of the number of correct classifications by the neural network learned without the given input variable, to the one obtained by the network trained with all input variables.

With the progress of pregnancy, the features characterizing the CTG signals change. In the monitoring system for a given date of CTG recording the current gestational age is determined in relation to the gestational age being entered to the database after calculation it with the aid of data obtained through the ultrasound examination that had been performed before the 20th week of gestation, confirmed by the modified Ballard method. That enabled two successive experiments to be performed in order to investigate how a different gestational age at which the signals were recorded, influences on the classification process.

In the first experiment the research material was partitioned taking into account the distribution of antenatal CTG traces recorded between 28th and 41st week of pregnancy. In order to ensure comparable number of records, four overlapping sets S1 ÷ S4 were proposed (Table 5). Additionally, 286 traces recorded at labour were extracted from the material and assigned to set SL. Classification was performed using the neural networks for each group separately. For every experiment by the trial and error method the best network structure was selected and basic parameters of the learning algorithm were set.

In the second the information on the gestational age was directly applied as an input in two ways: as the number of completed week of pregnancy, and as the number

of one of previously established four groups of antenatal traces. For a given gestational age expressed with accuracy of day its distances as an absolute differences to centres of the four groups were determined: G1 – 34.5 , G2 – 35.5, G3 – 36.5, G4 – 38.5. A given CTG record was assigned to the group with minimum distance. If the absolute difference was equal for two groups, this one of the higher centre was chosen. After conversion we obtained 228 traces with gestational age assigned to the G1, 105 to G2, 123 to G3 and 183 to G4 group.

Classifying the CTG signals as corresponding to abnormal or normal fetal state is a kind of diagnostic test giving positive or negative result, respectively. Relating it to a true result – the fetal outcome, allows for performance measurement using sensitivity (*SE*), specificity (*SP*), positive (*PPV*) and negative (*NPV*) predictive values. The evaluation of the neural networks performance is difficult when analysing all prognostic indices simultaneously. Therefore, the overall prognostic index *OI* was defined:

$$OI = \sqrt{\frac{(2 \cdot SE + NPV)}{3} \cdot \frac{(SP + PPV)}{2}} \quad [\%] \tag{5}$$

The sensitivity weight is doubled in *OI* formula because it is crucial to minimize the number of false negative cases, which they have more serious consequences than the false positive ones. We also calculated the percentage of correct classifications (*CC*).

3. Results and Discussion

An increase of number of epochs above 500 did not affect significantly the prediction quality. As for the learning rate the best results were obtained for 0.01 and 0.15. An increase of number of hidden neurons led to decrease of the generalization ability of the neural network and thus to decrease of the prediction quality for the testing set. The above tendencies were noted for all experiments. Table 2 presents three different classifiers, which provided the best results for particular combination of the data representation and the type of learning process applied. Testing the influence of data set representation on the neural networks efficiency we noticed decreasing of all mean values of prognostic indices for the categorical inputs (together with increasing of their SD values). In turn, changing the type of data learning from Real-Approach to Equal-Approach caused an increase of the *CC*, *SP* and *NPV* indices, whereas the *PPV* and the most important *SE* decreased. This tendency resulted in decreasing of the *OI*, but this change was not statistically significant. The results caused that in the next experiments the networks were designed with the numerical input data representation and the Real-Approach mode of learning.

Additionally, the importance index was calculated for all inputs, and then the inputs were ranked according to the index value. The mean values of ranks for a given

input in all trials, which represents its real influence on classification quality, are shown in Table 3. The rank value of 1 was assigned to the highest importance index, whereas 17 to the lowest one. It is easy to see, that the most significant parameter is the number of accelerations as well as the indices describing the FHR variability. This confirms that they are regarded as crucial signs of fetal wellbeing.

The influence of the gestational age, as an additional neural networks input, on the classification efficiency is presented in Table 4. The best classification quality was achieved for the MLP neural network with six hidden neurons fed with seventeen numerical parameters of CTG analysis. The additional input was the gestational age as a number of the gestational group, and the original proportion between normal and abnormal fetal outcomes was maintained. We obtained the sensitivity of 65.7% and the specificity of 68.5%.

Table 4. Classification results obtained for two NNs designed with two representations of gestational age applied as an additional input parameter

Gestational age representation	Completed week of pregnancy			Number of the antenatal group		
Prognostic index	Mean ± SD [%]	Min [%]	Max [%]	Mean ± SD [%]	Min [%]	Max [%]
SE	62.6 ± 8.0	45	78	65.7 ± 8.5	47	84
SP	64.4 ± 5.5	48	77	68.5 ± 6.2	53	79
PPV	42.2 ± 7.5	28	58	45.6 ± 7.9	25	61
NPV	80.6 ± 4.6	71	88	83.3 ± 4.6	75	93
CC	63.9 ± 6.6	51	78	67.7 ± 6.6	51	80
OI	60.5 ± 11.1	47	71	63.9 ± 10.2	47	74

Improvement of the classification quality was noted when the whole data set was divided into the five sets according to the gestational age and the neural networks were designed separately for each set (Table 5). In general, for the sets S1, S2 and S3 we obtained higher values for the most prognostic indices in comparison to the previous experiment, when gestational age was applied as an additional input. Taking into account statistical significance of the differences among the *OI* values, the best results were obtained for the neural network designed for the set S1. The highest sensitivity of 71.3% and specificity of 72.5% were obtained for the set of signals recorded in the earliest period of pregnancy, i.e. between 33rd and 36th week. With the increase of gestational age the decrease of prognostic indices was observed.
These results can be in some way related to those obtained in [22] – sensitivity of 73% and specificity of 94%, but it must be pointed out, that our results are much more rigorous because they concern mean values obtained after 50 trials with the randomly mixed learning and testing subsets. Additionally, in [22] much more input variables (30) were used and the abnormal fetal outcome was defined by more strict criteria applied to the newborn description.

Table 5. Characteristics of the sets with gestational and intrapartum traces and summary statistics of the associated classification results

Traces sets	SL	S1	S2	S3	S4
Gestational age [weeks]	Labour	<33÷36>	<34÷37>	<35÷38>	<36÷41>
Material characteristics	286 / 43%	285 / 32% [a]	284 / 31%	279 / 30%	278 / 29%
Prognostic index (Mean ± SD [%])					
SE	66.5 ± 8.4	71.3 ± 8.8	66.1 ± 10.3	67.8 ± 10.2	61.8 ± 10.1
SP	68.0 ± 9.7	72.5 ± 7.8	70.6 ± 7.4	67.0 ± 6.2	66.8 ± 8.2
PPV	62.0 ± 11.0	55.9 ± 10.3	51.4 ± 10.0	46.6 ± 7.7	43.1 ± 10.4
NPV	72.2 ± 7.5	83.8 ± 5.9	81.7 ± 6.1	83.1 ± 5.5	81.6 ± 4.7
CC	67.4 ± 9.6	72.1 ± 7.8	69.2 ± 8.2	67.3 ± 7.4	65.5 ± 9.8
OI	66.6 ± 13.2 [b]	**69.6 ± 13.4** [c]	66.0 ± 13.0	64.3 ± 12.6	61.3 ± 12.9
Prognostic index (Min ÷ Max)					
SE	45 ÷ 90	50 ÷ 90	42 ÷ 95	43 ÷ 87	35 ÷ 83
SP	44 ÷ 89	52 ÷ 88	54 ÷ 84	48 ÷ 77	48 ÷ 83
PPV	35 ÷ 85	39 ÷ 78	30 ÷ 68	24 ÷ 63	22 ÷ 65
NPV	56 ÷ 90	69 ÷ 94	67 ÷ 97	70 ÷ 93	71 ÷ 91
CC	46 ÷ 75	59 ÷ 82	57 ÷ 77	50 ÷ 83	51 ÷ 71
OI	45 ÷ 78	57 ÷ 83	50 ÷ 78	48 ÷ 77	47 ÷ 75

[a] number of records / percentage with abnormal fetal outcome.

[b] statistically significant difference SL vs. S4 ($p < 0.01$).

[c] statistically significant difference S1 vs. S2, S3, S4, SL ($p < 0.02$).

4. Conclusions

In the presented work, a number of experiments were done in order to show an influence of the input data on classification of CTG signals as predicting normal or abnormal fetal outcome using the MLP neural networks. Classification was carried out through seventeen selected parameters of the computerized quantitative analysis of CTG record. These parameters were fed to the neural networks input layer in the original numerical form, as well as in the categorical form after their conversion basing on established ranges of physiology. The real fetal outcome was defined as abnormal, if the value of at least one attribute (Apgar score or percentile of birth weight or pH level) was outside the physiological range.

Various structures of learning subsets were tested to consider that during pregnancy the fetus is usually monitored several times which leads to assigning a number of CTG traces to one fetal outcome. Representation of the input variables in the categorical form caused a decrease of all performance indices (with an increase of their standard deviations), so the numerical representation should be preserved. However, it could be connected with the fact, that the number of categorical input variables (11) was smaller than the number of numerical ones (17).

Among the numerical input parameters the most significant were the number of accelerations and the indices describing the instantaneous variability of the FHR. Improvement of the classification quality was noted when whole data set was divided into the sets according to the gestational age, and the neural networks were designed separately for each set. However, the classification quality indices decreased with increase of gestational age. We found these results as encouraging and the plan for our future research is to collect larger database in order to select more representative groups with gestational as well as intrapartum traces.

Acknowledgments

This work was supported in part by the Ministry of Sciences and Higher Education resources in 2010–2012 under the Research Project N N518 411138.

References

1. Jeżewski J., Wrobel J., Horoba K., Kupka T., Matonia A.: Centralised fetal monitoring system with hardware-based data flow control. Proc. III Inter. Conf. MEDSIP, 2006, 51–54.
2. Jeżewski M., Wrobel J., Horoba K., Gacek A., Henzel N., Leski J.: The prediction of fetal outcome by applying neural network for evaluation of CTG records. In: Computer Recognition Systems, Kurzynski M., Puchala E. (Eds), Advances in Soft Computing Series, Springer Verlag, 2007, 45, 532–541.
3. Catley C., Frize M., Walker R., Petriu D.: Predicting high-risk preterm birth using artificial neural networks. IEEE Trans. Inf. Technol. Biomed. 2006, 10, 540–549.
4. Jeżewski M., Czabanski R., Roj D., Wrobel J., Jeżewski J.: Application of RBF neural networks for predicting low birth weight using features extracted from fetal monitoring signals. IFMBE Proc. World Congr. Med. Phys. Biomed. Eng. 2009, 25, 374–377.
5. Jeżewski M., Czabanski R., Horoba K., Wrobel J.: Prediction of newborn sex with neural networks approach to fetal cardiotocograms classification. In: Information Technologies in Biomedicine, Pietka E., Kawa J. (Eds). Advances in Soft Computing Series, Springer Verlag, 2008, 47, 299–306.
6. Beksac M., Ozdemir K., Erkmen A., Karakas U.: Assessment of antepartum fetal heart rate tracings using neural networks. In: A critical appraisal of fetal surveillance. Van Geijn H.P., Copray F.J.(Eds), Elsevier Science B.V, 1994, 354–362.
7. Georgoulas G., Stylios C., Groumpos P.: Predicting the risk of metabolic acidosis fornewborns based on fetal heart rate signal classification using support vector machines, IEEE Trans. Biomed. Eng. 2006, 53, 875–884.
8. Arduini D., Giannini F., Magnes G., Signorini M.G.: Fuzzy logic in the management of new perinatal variables. Proc. 5th World Congr. Perinatal Medicine, Carrera J.M., Monduzzi E. (Eds), 2001, 1211–1216.
9. Salamalekis E., Thomopoulos P., Giannaris D., Salloum I.: Computerised intrapartum diagnosis of fetal hypoxia based on fetal heart rate monitoring and fetal pulse oximetry recordings utilising wavelet analysis and neural networks. Int. J. Obstet. Gynaecol. 2002, 109, 1137–1142.
10. Georgoulas G., Gavrilis D., Tsoulos J.G., Stylios C.: Novel approach for fetal heart rate classification introducing grammatical evolution. Biomed. Signal Proc. Control 2007, 2, 69–79.
11. Georgoulas G., Stylios C., Groumpos P.: Feature extraction and classification of fetal heart rate using wavelet analysis and support vector machines. Int. J. Artif. Intel. Tools 2006, 15, 411–432.

12. Magenes G., Signorini M.G., Sassi R.: Multiparametric analysis of fetal heart rate: comparison of neural and statistical classifiers. Proc. 9th Conf. Med. Eng. Comput., 2001, 360–363.
13. Hasbargen U.: Application of neural networks for intrapartum surveillance. In: A critical appraisal of fetal surveillance, Geijn H.P., Copray F.J. (Eds), Elsevier Science BV 1994, 363–367.
14. Fontenla-Romero O., Alonso-Betanzos A., Guijarro-Berdinas B.: Adaptive pattern recognition in the analysis of cardiotocographic records. IEEE Trans. Neural Networks 2001, 12, 1188–1195.
15. Guijarro-Berdinas B., Alonso-Betanzos A., Fontenla-Romero O.: Intelligent analysis and pattern recognition in cardiotocographic signals using a tightly coupled hybrid system, Artif. Intel. 2002, 136, 1–27.
16 Keith R., Westgate J., Ifeachor E., Greene K.: Suitability of artificial neural networks for feature extraction during labour. Med. Biol. Eng. Comput.1994, 32, 51–57.
17 Kol S., Thaler I., Paz N., Shmueli O.: Interpretation of nonstress test by an artificial neural networks. Am. J. Obstet. Gynecol. 1995, 172, 1372–1378.
18. Magenes G., Signorini M.G., Sassi R.: Automatic diagnosis of fetal heart rate: comparison of different methodological approaches. Proc. 23rd Intern. Conf. IEEE EMBS 2001, 1, 1604–1607.
19. Czabanski R., Jeżewski M., Wróbel J., Jeżewski J.: Fuzzy system for evaluation of fetal heart rate signals using FIGO criteria. J. Med. Inform. Technol. 2009, 13, 189–194.
20. Czabanski R., Jeżewski M., Wróbel J., Jeżewski J., Horoba K.: Predicting the risk of low fetal birth weight from cardiotocographic signals using ANBLIR system with deterministic annealing and e-insensitive learning. IEEE Trans. Inf. Technol. BioMed. 2010, 14, 1062–1074.
21. Magenes G., Signorini M.G., Arduini D.: Classification of cardiotocographic records by neural networks. Proc. IEEE Intern. Joint Conf. Neural Networks 2000, 3, 637–641.
22. Arduini D., Giannini F., Magenes G.: Intrapartum surveillance – computer cardiotocography. Proc. 5th World Congress of Perinatal Medicine, 2001, 1217–1223.
23. Ayres-de-Campos D., Costa-Santos C., Bernardes J.: Prediction of neonatal state by computer analysis of fetal heart rate tracings – the antepartum arm of the SisPorto multicentre validation study. Eur. J. Obstet. Gynecol. Reprod. Biol. 2005, 118, 52–60.
24. Dawes G.S., Moulden M., Redman C.W.: Limitations of antenatal fetal heart rate monitors. Am. J. Obstet. Gynecol. 1990, 162, 170–173.
25. Jeżewski J., Kupka T., Horoba K.: Extraction of fetal heart rate signal as time event series from evenly sampled data acquired using Doppler ultrasound technique. IEEE Trans. Biomed. Eng. 2008, 55, 805–810.
26. van Geijn H.P.: Analysis of heart rate and beat-to-beat variability: Interval difference index. Am. J. Obstet. Gynecol. 1980, 138, 246–252.
27. Jeżewski J., Wrobel J., Horoba K., Moczko J.: Advances in Doppler ultrasound FHR monitoring. Klin. Perinat. Ginekol. 1995, 3, 241–251.
28. Rooth G.: Guidelines for the use of fetal monitoring. Int. J. Gynecol. Obstet. 1987, 25, 159–167.
29. Dawes G.S., Houghton C.R.: Baseline in human fetal heart rate records. Br. J. Obstet. Gynaecol. 1982, 89, 270–275.
30. Mantel R., van Geijn H.P., Caron F.J., Swartjes J.M.: Computer analysis of antepartum fetal heart rate: 1. Baseline determination. Int. J. Bio. Med. Comput. 1990, 25, 261–272.
31. Jeżewski J., Wrobel J., Horoba K.: Comparison of Doppler ultrasound and direct electrocardiography acquisition techniques for quantification of fetal heart rate variability. IEEE Trans. Biomed. Eng. 2006, 53, 855–864.
32 Jeżewski J., Horoba K., Matonia A., Wrobel J.: Quantitative analysis of contraction patterns in electrical activity signal of pregnant uterus as an alternative to mechanical approach. Physiol, Meas. 2005, 26, 753–767.
33. Jeżewski M., Henzel N., Wrobel J., Labaj P.: Application of neural networks for prediction of fetal outcome. J. Med. Inform. Technol. 2006, 10, 127–132.
34. Sikora J.: Digital analysis of cardiotocographic traces for clinical fetal outcome prediction. Klin. Perinat. Ginekol. 2001, 10, 57–88.